

Comparing machine learning models to choose the variable ordering for cylindrical algebraic decomposition

Matthew England and Dorian Florescu

Faculty of Engineering, Environment and Computing,
Coventry University, Coventry CV1 5FB, UK
{Matthew.England, Dorian.Florescu}@coventry.ac.uk

Abstract. There has been recent interest in the use of machine learning (ML) approaches within mathematical software to make choices that impact on the computing performance without affecting the mathematical correctness of the result. We address the problem of selecting the variable ordering for cylindrical algebraic decomposition (CAD), an important algorithm in Symbolic Computation. Prior work to apply ML on this problem implemented a Support Vector Machine (SVM) to select between three existing human-made heuristics, which did better than anyone heuristic alone. Here we extend this result by training ML models to select the variable ordering directly, and by trying out a wider variety of ML techniques.

We experimented with the NLSAT dataset and the Regular Chains Library CAD function for Maple 2018. For each problem, the variable ordering leading to the shortest computing time was selected as the target class for ML. Features were generated from the polynomial input and used to train the following ML models: k-nearest neighbours (KNN) classifier, multi-layer perceptron (MLP), decision tree (DT) and SVM, as implemented in the Python scikit-learn package. We also compared these with the two leading human-made heuristics for the problem: the Brown heuristic and *sotd*. On this dataset all of the ML approaches outperformed the human-made heuristics, some by a large margin.

Keywords: computer algebra; symbolic computation; non-linear real arithmetic; cylindrical algebraic decomposition; machine learning.

1 Introduction

A logical statement is *quantified* if it involves the universal quantifier \forall or the existential quantifier \exists . The *Quantifier Elimination* (QE) problem is to derive from a quantified formula an equivalent un-quantified one. A simple example would be that the quantified statement, “ $\exists x. x^2 + bx + c = 0$ ” is equivalent to the unquantified statement “ $b^2 - 4c \geq 0$ ”, when working over the real numbers. QE is one definition for simplifying or solving a problem. The tools involved fall within the field of Symbolic Computation, implemented in Computer Algebra Systems (or more recently in SMT-solvers).

Our work is on Quantifier Elimination over the reals. Here the logical statements are expressed as *Tarski formulae*, Boolean combinations ($\wedge, \vee, \neg, \rightarrow$) of statements about the signs of polynomials with integer coefficients. QE in this theory was first shown to be soluble by Tarski [53] in the 1940s. However, the only implemented general real QE procedure has algorithmic complexity doubly exponential in the number of variables [22], a theoretical result experienced in practice. For many problem classes QE procedures will work well at first, but as the problem size increases the doubly exponential wall is inevitably hit. It is hence of critical importance to optimise QE procedures and the formulation of problems, to “push the doubly exponential wall back” and open up a wider range of tractable applications.

QE procedures can be run in multiple ways to solve a given problem: they can be initialized with different options (e.g. variable ordering [23], equational constraint designation [10]); tasks can be completed in different orders (e.g. order of constraint analysis [26]); and the problem itself may be expressible in different formalisations [56]. Changing these settings can have a substantial effect on the computational costs (both time and memory) but does not effect the mathematical correctness of the output. They are thus suitable candidates for machine learning: tools that allow computers to make decisions that are not explicitly programmed, via the statistical analysis of large quantities of data.

We continue in Section 2 by introducing background material on the particular decision we study: the variable ordering for Cylindrical Algebraic Decomposition. Here we also outline prior attempts to solve this problem; and prior applications of machine learning to computer algebra. Then in Section 3 we describe our methodology covering datasets, software, features extracted from the problems, machine learning models tests, and how we test against human-made heuristics. We present our results in Section 4 and final thoughts in Section 5.

2 Variable Ordering for CAD

2.1 Cylindrical algebraic decomposition

A *Cylindrical Algebraic Decomposition* (CAD) is a *decomposition* of ordered \mathbb{R}^n space into cells arranged *cylindrically*: meaning the projections of any pair of cells with respect to the variable ordering are either equal or disjoint. The cells are (semi)-algebraic meaning each cell can be described with a finite sequence of polynomial constraints. A CAD is produced to be invariant relative to an input, i.e. *truth-invariant* for a logical formula (so the formula is either true or false throughout each cell). Such a decomposition can then be used to perform quantifier elimination on the formula by testing a finite set of sample points and constructing a quantifier-free formula from the semi-algebraic cell descriptions.

CAD was introduced by Collins in 1975 [19] and works relative to a set of polynomials. Collins’ CAD produces a decomposition so that each polynomial has constant sign on each cell (thus truth invariant for any formula built with those polynomials). The algorithm first projects the polynomials into smaller

and smaller dimensions; and then uses these to lift — to incrementally build decompositions of larger and larger spaces according to the polynomials at that level. For full details on the original CAD algorithm see [3].

QE has numerous applications throughout science and engineering (see for example the survey [52]). New applications are found regularly, such as the derivation of optimal numerical schemes [31], and the validation of economic hypotheses [46], [45]. CAD has also found application independent of QE, such as reasoning with multi-valued functions [21] (where we decompose to see where simplification rules are valid); and biological networks [7], [29] where we decompose to identify regions in parameter space where multi-stationarity can occur.

The definition of cylindricity and both stages of the algorithm are relative to an ordering of the variables. For example, given polynomials in variables ordered as $x_n \succ x_{n-1} \succ \dots \succ x_2 \succ x_1$ we first project away x_n and so on until we are left with polynomials univariate in x_1 . We then start lifting by decomposing the x_1 -axis, and then the (x_1, x_2) -plane and so so on. The cylindricity condition refers to projections of cells in \mathbb{R}^n onto a space (x_1, \dots, x_m) where $m < n$.

There have been numerous advances to CAD since its inception, for example: on how best to implement the projection [34], [43], [12], [44]; avoiding the need for full CAD [20], [55]; symbolic-numeric lifting schemes [50], [38]; adapting to the Boolean structure in the input [8], [9], [27]; and local projection approaches [15], [51]. However, in all cases, the need for a fixed variable ordering remains.

2.2 Effect of the variable ordering

Depending on the application requirements the variable ordering may be determined, constrained, or entirely free. The most common application, QE, requires that the variables be eliminated in the order in which they are quantified in the formula but makes no requirement on the free variables. For example, we could eliminate the quantifier in $\exists x. ax^2 + bx + c = 0$ using any CAD which eliminates x first; giving six possible orderings to choose from. A CAD for the polynomial under ordering $a \prec b \prec c$ has only 27 cells, but needs 115 for the reverse ordering.

Note that since we can switch the order of quantified variables in a statement when the quantifier is the same, we also have some choice on the ordering of quantified variables. For example, a QE problem of the form $\exists x \exists y \forall a \phi(x, y, a)$ could be solved by a CAD under either ordering $x \succ y \succ a$ or ordering $y \succ x \succ a$.

The choice of variable ordering has been long known to have a great effect on the time and memory use of CAD, and the number of cells in the output. In fact, Brown and Davenport presented a class of problems in which one variable ordering gave output of double exponential complexity in the number of variables and another output of a constant size [14].

2.3 Prior work on choosing the variable ordering

Heuristics have been developed to choose a variable ordering, with Dolzmann et al. [23] giving the best known study. After analysing a variety of metrics they

proposed a polynomial degree based heuristic (the heuristic *sotd* defined later). However the second author demonstrated examples for which that heuristic could be misled in CICM 2013 [10]; showed that tailoring it to an implementation could improve its performance in ICMS 2014 [28]; and in CICM 2014 [37] reported that a computationally cheaper heuristic by Brown actually outperforms *sotd*.

In CICM 2014 [37] we used a support vector machine (SVM), an ML model widely used for non-linear classification, to choose which of three human-made heuristics to believe when picking the variable ordering. The experiments in [37] identified substantial subclasses on which each of the three heuristics made the best decision, and demonstrated that the machine learned choice did significantly better than any one heuristic overall. This motivated the present study where we consider a wider range of machine learning models and have these pick the ordering directly from the full range of choices.

2.4 Other applications of ML to mathematical software

The CICM 2014 paper was the first to document the application of machine learning to CAD, or in fact to any symbolic computation algorithm / computer algebra system. Since then there have been two further studies:

- The same authors studied a different choice related to CAD (whether to precondition the input with Gröbner bases) in [35] [36], again finding that a support vector machine could make the choice more accurately than the human-made heuristic (if features of the Gröbner Basis could be used).
- At MACIS 2016 there was a study applying a support vector machine to decide the order of sub-formulae solving for a QE procedure [40].

The survey paper [25] and the ICMS Special Session on Machine Learning for Mathematical Software demonstrated the wide range of other potential applications. As discussed there, while the use of machine learning in computer algebra is rare it has become a key tool in other mathematical software development. Most notably automated reasoning [54], [41], [11], [2]; but also satisfiability checking [59], [42].

3 Methodology

3.1 Dataset

Despite its long history and significant software contributions the Computer Algebra community had a lack of substantial datasets [33]: a significant barrier to machine learning. Despite efforts to address this¹, the most substantial dataset of problems designed for CAD is [58] with less than 100 examples.

However, CAD has recently found prominence in a new area: Satisfiability Modulo Theories (SMT). Here, efficient algorithms for the Boolean SAT problem

¹ E.g. the PoSSo and FRISCO projects in the 90s and the SymbolicData Project [32].

study the Boolean skeleton of a problem, with a theory solver then checking if a satisfying Boolean assignment is valid in the domain (learning new clauses if not) [4]. For the SMT domain of Non-linear Real Arithmetic (NRA), CAD can play the role of such theory solvers², and so their test problems may be used to evaluate CAD. We use the `nlsat` dataset³ produced to evaluate the work in [39]. The main sources of the examples are: METiTARSKI [47], an automatic theorem prover for theorems with real-valued special functions (it applies real polynomial bounds and then using QE tools like CAD); problems originating from attempts to prove termination of term-rewrite systems; verification conditions from Keymaera [49]; and parametrized generalizations of geometric problems.

The problems in the `nlsat` dataset are all fully existential (the only quantifier is \exists) which is why they may be studied by SAT solvers. Although CAD can make adaptations based on the quantifiers in the input (most notably via Partial CAD [20]) the conclusions drawn are likely to be applicable outside of the SAT context.

We extracted 6117 problems with 3 variables from this database, meaning each has a choice of six different variable orderings. We randomly divided them into two datasets for training (4612) and testing (1505). The training dataset was used to tune the parameters of the ML models. The testing dataset was unknown to the models during training, and is used to compare the performance of the different ML models and the human-made heuristics.

3.2 Software

We experimented using the CAD routine `CylindricalAlgebraicDecompose` which is part of the `RegularChains` Library for MAPLE. This algorithm builds decompositions first of n -dimensional complex space before refining to a CAD of \mathbb{R}^n [18], [17], [6]. We ran the code in Maple 2018 but used an updated version of the `RegularChains` Library downloaded from <http://www.regularchains.org>, which contains bug fixes and additional functionality. We ignored the quantifiers and logical connectives, using only the polynomials as input to CAD. The function thus returned a sign-invariant CAD for the polynomials.

The training and evaluation of the machine learning models was done using the `scikit-learn` package [48] v0.20.2 for Python 2.7. The features for machine learning were extracted using code written in the `sympy` package v1.3 for Python 2.7, as was the Brown heuristic. The `sotd` heuristic was implemented in MAPLE as part of the `ProjectionCAD` package [30].

3.3 Timings

CAD construction was timed in a Maple script that was called using the `os` package in Python for each CAD, to avoid Maple caching of results.

The target variable ordering for ML was defined as the one that minimises the computing time for a given SAT problem. All CAD function calls included a

² However, as discussed by [1] a more custom approach is beneficial.

³ <http://cs.nyu.edu/~dejan/nonlinear/>

time limit. The problems in the training dataset were processed with an initial time limit of 4 seconds on all variable orderings. The time limit was doubled if all 6 orderings timed out. The process stopped when the CAD routine was completed for at least one ordering. All problems in the training dataset could be assigned a target variable ordering using time limits smaller than 64 seconds.

The problems in the testing dataset were processed with a larger time limit of 128 seconds for all orderings. This was in order to allow a better comparison of the computing times for the ML and the heuristics. When a variable ordering timed out, the computing time was considered equal to 128 seconds.

3.4 Features

We computed the same set of 11 features for each SAT problem as [37], which are listed in Table 1. All these features could be computed from the input polynomials immediately. A possibility for future work is to consider features that are more expensive, such as those from post-processing as in [35], those from the end of CAD projection as sotd does, or perhaps even going further into partial lifting as in [57]. The ML models associate a predicted variable ordering to each set of 11 features. The training and testing feature set were normalised using the mean and standard deviation of the training set.

Table 1. The features used by ML to predict variable orderings.

| Feature number | Description |
|----------------|--|
| 1 | Number of polynomials. |
| 2 | Maximum total degree of polynomials. |
| 3 | Maximum degree of x_0 among all polynomials. |
| 4 | Maximum degree of x_1 among all polynomials. |
| 5 | Maximum degree of x_2 among all polynomials. |
| 6 | Proportion of x_0 occuring in polynomials. |
| 7 | Proportion of x_1 occuring in polynomials. |
| 8 | Proportion of x_2 occuring in polynomials. |
| 9 | Proportion of x_0 occuring in monomials. |
| 10 | Proportion of x_1 occuring in monomials. |
| 11 | Proportion of x_2 occuring in monomials. |

3.5 ML models

Four of the most commonly used deterministic ML models were tuned on the training data (for details on the methods see for example the textbook [5]):

- The K-Nearest Neighbours (KNN) classifier [5, §2.5].
- The Multi-Layer Perceptron (MLP) classifier [5, §2.5].
- The Decision Tree (DT) classifier [5, §14.4].
- The Support Vector Machine (SVM) classifier with RBF kernel [5, §6.3].

The KNN classifier is a type of *instance-based* classifier, i.e. it does not construct an internal model of the data but stores all the instances in the training data for prediction. So for each new data instance the model selects the nearest k training instances. Selection can be performed by weighting instances equally, or by weighting a training instance inversely proportional to its distance from the new instance. The prediction is given by the class with the highest count among those training instances. Three algorithms are typically used to train KNN: the *brute force* algorithm computes the distances between all pairs of points, the *k-dimensional tree* algorithm partitions the data along Cartesian axes, and the *ball tree* algorithm partitions the data in a series of nesting hyper-spheres.

The DT is a non-parametric model that uses a tree-like model of decisions and their possible consequences. Each node in the tree represents a test on an attribute, each branch the outcome of the test. The *leaves* are the end points of each branch, representing the predicted class label. There are two common criteria used to assess the quality of a split in the DT. The Gini impurity criterion verifies how often a randomly chosen element would be correctly labelled if it were randomly labeled according to the distribution of labels in the subset. The entropy criterion assesses the information gain after each split.

The SVM is a model that can perform linear and non-linear classification, by selecting an appropriate kernel. It is also known as a maximum-margin classifier, because it identifies a hyperplane in the feature space that maximises the distance to the nearest data points⁵. The SVM kernel acts as a similarity function between the training examples and a few predefined *landmarks*, that can offer additional computing performance. The most common kernels are: *linear*, *polynomial*, *sigmoidal* and *radial basis function* (RBF). The RBF kernel is one of the most common kernel choices, given by

$$\mathcal{K}(\mathbf{x}, \ell) = e^{-\gamma \cdot \|\mathbf{x} - \ell\|^2},$$

where γ is the kernel hyperparameter, and $\ell \in \mathbb{R}^n$ is a predefined *landmark*, and $\mathbf{x} \in \mathbb{R}^n$ is the training vector with n features.

The MLP is a class of feedforward artificial neural networks. It consists of a minimum of 3 layers: the input, hidden and output layer. Both the hidden and output layers use a nonlinear activation function that can be selected during cross-validation. Some of the common activation functions are: *identity*, *logistic*, *hyperbolic tangent*, and *rectified linear*.

3.6 Training

Each model was trained using grid search 5-fold cross-validation. Specifically, the training feature set was randomly divided in 5 equal parts. Each of the possible combinations of 4 parts was used to tune the model parameters, leaving the last part for fitting the hyperparameters by cross-validation. For each of the models, the grid search was performed for an initially large range for each hyperparameter. This range was increased until all optimal hyperparameters were inside the range, and not on the edge. Subsequently, the range was gradually decreased

Table 2. The hyperparameters of the ML models optimised with 5-fold cross-validation on the training dataset.

| Model | Hyperparameter | Value |
|------------------------|--|---|
| Decision Tree | Criterion | Gini impurity |
| | Maximum tree depth | 17 |
| K-Nearest Neighbours | Train instances weighting Algorithm | Inversely proportional to distance Ball Tree |
| Support Vector Machine | Regularization parameter C | 316 |
| | Kernel | Radial basis function |
| | γ | 0.08 |
| Multi-Layer Perceptron | Tolerance for stopping criterion | 0.0316 |
| | Hidden layer size | 18 |
| | Activation function | Hyperbolic tangent |
| | Algorithm | Quasi-Newton based optimiser |
| | Regularization parameter α | $5 \cdot 10^{-5}$ |

to home in on each optimal hyperparameter, until the performance plateaued. Each grid search lasted from a few seconds for simpler models like KNN to a few minutes for more complex models like MLP. The optimal hyperparameters selected during cross-validation are in Table 2.

3.7 Comparing with human-made heuristics

The ML approaches were compared in terms of prediction accuracy and resulting CAD computing time against the two best known human-made heuristics.

Brown This heuristic chooses a variable ordering according to the following criteria, starting with the first and breaking ties with successive ones:

- (1) Eliminate a variable first if it has lower overall degree in the input.
- (2) Eliminate a variable first if it has lower (maximum) total degree of those terms in the input in which it occurs.
- (3) Eliminate a variable first if there is a smaller number of terms in the input which contain the variable.

It is named after Brown who documented it only in the notes of an ISSAC⁴.

sotd This heuristic constructs the full set of projection polynomials for each permitted ordering and selects the ordering whose corresponding set has the lowest **sum of total degrees** for each of the monomials in each of the polynomials. It was the recommendation made after the study [23].

Unlike the ML, these human-made heuristics can end up predicting several variable orderings (i.e. when they cannot discriminate). In practice if this were to happen the heuristic would select one randomly (or perhaps lexicographically), however that final pick is not particularly meaningful for an evaluation. To accommodate this, for each problem, the prediction accuracy of such a heuristic is

⁴ <https://www.usna.edu/Users/cs/wcbrown/research/ISSAC04/handout.pdf>

judged to be the the percentage of its predicted variable orderings that are also target orderings. The average of this percentage over all problems in the testing dataset represents the prediction accuracy. Similarly, the computing time for such methods was assessed as the average computing time over all predicted orderings, and it is this that is summed up for all problems in the testing dataset.

4 Results

We compare the four ML models on the percentage of problems where they selected the optimum ordering, and the total computation time (in seconds) for solving all the problems with their chosen orderings. We also compare the ML models with the two human-made heuristics (with the adaptations outlined in Section 3.7) and finally the outcome of a random choice between the 6 orderings. The results are presented in Table 3. We might expect a random choice to be correct one sixth of the time for this data set (16.6%). The actual accuracy is a little higher because the dataset is not uniform, and for some problem instances there were multiple variable orderings with equally fast timings.

Moreover, we evaluate the distribution of the computation time for the ML methods and the heuristics. The differences between the computation time of each method and the minimum computation time, given as a percentage of the minimum time, are depicted in Figure 1.

4.1 Range of possible outcomes

We note that the minimum total computing time achieved by selecting an optimal variable ordering for every problem would be 8,623s. Choosing at random would take 30,235s, almost 4 times as much. The maximum total computing time, determined by selecting the variable ordering with the longest computing time, is 64,534s. The choices with the quickest time among the methods considered were achieved by the Decision Tree model: 9,994s, which is 16% more than the minimal possible. So there are clearly great time savings to be made by taking this choice into account.

4.2 Human-made heuristics

Of the two human-made heuristics, Brown performed the best, as it did in [37]. As was noted there this is surprising since the sotd heuristic has access to ad-

Table 3. The comparative performance of DT, KNN, MLP, SVM, and the Brown and sotd heuristics on the testing dataset.

| | DT | KNN | MLP | SVM | Brown | sotd | random |
|-----------------------------|-------|--------|-------|--------|--------|--------|--------|
| Accuracy | 62.6% | 63.3% | 61.6% | 58.8% | 51% | 49.5% | 22.7% |
| Computation Time (s) | 9,994 | 10,105 | 9,822 | 10,725 | 10,951 | 11,938 | 30,235 |

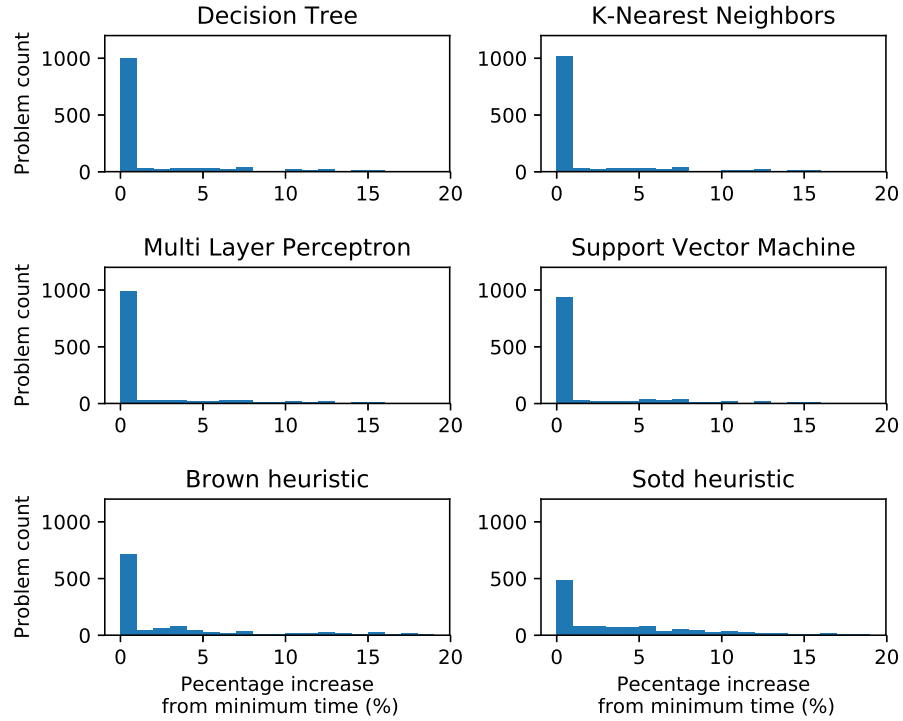


Fig. 1. The histograms of the percentage increase in computation time relative to the minimum computation time for each method, calculated for a bin size of 1%.

ditional information (not just the input polynomials but also their projections). Unlike the ML models and the Brown heuristic, obtaining an ordering for a problem instance with sotd is not instantaneous. Generating an ordering with sotd for all problems in the testing dataset took over 30min.

Brown could solve all problems in 10,951s, 27% more than the minimum. While sotd is only 0.7% less accurate than Brown in identifying the best ordering, it is much slower at 11,938s or 38% more than the minimum. This shows that while Brown is not much better at identifying the best ordering, it does a much better job at discarding the worst!

4.3 ML choices

The results show that all ML approaches outperform the human-made heuristics in terms of both accuracy and timings. Figure 1 shows that the human-made heuristics result in computing times that are often significantly larger than 1% of the corresponding minimum time for each problem. The ML methods, on the

other hand, all result in almost 1000 problems ($\sim 75\%$ of the testing dataset) within 1% of the minimum time.

The key finding of the present paper is that there are significantly better models to use for the problem than SVM: each of the other three had 10% higher accuracy and much lower timings. It is interesting to note that the MLP model leads to a lower accuracy than DT and KNN, but achieves the lowest overall computing time. This suggests that on this dataset, the MLP model is best at leaving out variable orderings that lead to long computing times, even if it has a slightly higher chance of missing the target variable ordering.

4.4 Conclusion

For a casual user of computer algebra the key consideration is probably that their computation does not hang, and that it finishes reasonably fast on the widest variety of questions. Hence of the tested models we may recommend the MLP for use going forward. However, all the models show promise and we emphasise that this conclusion applies only for the present dataset and we need wider experimentation to see if this finding is replicated.

5 Final Thoughts

The experiment shows a clear advance on the prior state of the art: both in terms of accuracy of predictions for CAD variable ordering; and in understanding which models are the best suited for this problem. But we acknowledge that there is much more to do and emphasise that this is only the initial findings of the project. The following extensions will be undertaken shortly:

- Extending the experiment to include problems with more variables from the dataset. Unlike some CAD implementations the one used does not change algorithm when $n > 3$; however, like all CAD implementation the time required will increase exponentially.
- Using the trained classifiers to test on CAD problems from outside the dataset, in particular, those which are not fully existentially quantified (SAT problems). In such cases the algorithms can change (terminate early) and it is not clear if models trained for the SAT case can be applied there.
- Using the trained classifiers to test on CAD implementations other than the one in the MAPLE `RegularChains` Library. For example, can the classifiers also usually pick variable orderings for QEPCAD-B, [13], or REDLOG [24].
- Examining how best to derive additional features to use for the training, and the use of feature selections tools to find an optimal subset.
- Test a dynamical selection of the variable ordering, where ML only picks the first variable for a problem, the polynomials are projected along that variable, and then the process repeats iteratively.

Finally, we note the wide variety of additional problems in computer algebra systems open to study with machine learning [25].

Acknowledgements

The authors are supported by EPSRC Project EP/R019622/1: *Embedding Machine Learning within Quantifier Elimination Procedures*.

Research Data Statement Data supporting the research in this paper is available at: <http://doi.org/10.5281/zenodo.2658626>.

References

1. Ábrahám, E., Abbott, J., Becker, B., Bigatti, A., Brain, M., Buchberger, B., Cimatti, A., Davenport, J., England, M., Fontaine, P., Forrest, S., Griggio, A., Kroening, D., Seiler, W., Sturm, T.: SC^2 : Satisfiability checking meets symbolic computation. In: Kohlhase, M., Johansson, M., Miller, B., de Moura, L., Tompa, F. (eds.) *Intelligent Computer Mathematics: Proc. CICM '16*, Lecture Notes in Computer Science, vol. 9791, pp. 28–43. Springer International Publishing (2016), https://doi.org/10.1007/978-3-319-42547-4_3
2. Alemi, A., Chollet, F., Een, N., Irving, G., Szegedy, C., Urban, J.: Deepmath – Deep sequence models for premise selection. In: *Proc. 30th International Conference on Neural Information Processing Systems*. pp. 2243–2251. NIPS '16, Curran Associates Inc. (2016), <https://papers.nips.cc/paper/6280-deepmath-deep-sequence-models-for-premise-selection.pdf>
3. Arnon, D., Collins, G., McCallum, S.: Cylindrical algebraic decomposition I: The basic algorithm. *SIAM Journal of Computing* **13**, 865–877 (1984), <https://doi.org/10.1137/0213054>
4. Barrett, C., Sebastiani, R., Seshia, S., Tinelli, C.: Satisfiability modulo theories. In: Biere, A., Heule, M., van Maaren, H., Walsh, T. (eds.) *Handbook of Satisfiability* (Volume 185 *Frontiers in Artificial Intelligence and Applications*), Chapter 26, pp. 825–885. IOS Press (2009)
5. Bishop, C.: *Pattern Recognition and Machine Learning*. Springer (2006)
6. Bradford, R., Chen, C., Davenport, J., England, M., Moreno Maza, M., Wilson, D.: Truth table invariant cylindrical algebraic decomposition by regular chains. In: Gerdt, V., Koepf, W., Seiler, W., Vorozhtsov, E. (eds.) *Computer Algebra in Scientific Computing, Lecture Notes in Computer Science*, vol. 8660, pp. 44–58. Springer International Publishing (2014), http://dx.doi.org/10.1007/978-3-319-10515-4_4
7. Bradford, R., Davenport, J., England, M., Errami, H., Gerdt, V., Grigoriev, D., Hoyt, C., Košta, M., Radulescu, O., Sturm, T., Weber, A.: A case study on the parametric occurrence of multiple steady states. In: *Proc. 42nd International Symposium on Symbolic and Algebraic Computation*. pp. 45–52. ISSAC '17, ACM (2017), <https://doi.org/10.1145/3087604.3087622>
8. Bradford, R., Davenport, J., England, M., McCallum, S., Wilson, D.: Cylindrical algebraic decompositions for boolean combinations. In: *Proc. 38th International Symposium on Symbolic and Algebraic Computation*. pp. 125–132. ISSAC '13, ACM (2013), <http://dx.doi.org/10.1145/2465506.2465516>
9. Bradford, R., Davenport, J., England, M., McCallum, S., Wilson, D.: Truth table invariant cylindrical algebraic decomposition. *Journal of Symbolic Computation* **76**, 1–35 (2016), <http://dx.doi.org/10.1016/j.jsc.2015.11.002>

10. Bradford, R., Davenport, J., England, M., Wilson, D.: Optimising problem formulations for cylindrical algebraic decomposition. In: Carette, J., Aspinall, D., Lange, C., Sojka, P., Windsteiger, W. (eds.) *Intelligent Computer Mathematics, Lecture Notes in Computer Science*, vol. 7961, pp. 19–34. Springer Berlin Heidelberg (2013), http://dx.doi.org/10.1007/978-3-642-39320-4_2
11. Bridge, J., Holden, S., Paulson, L.: Machine learning for first-order theorem proving. *Journal of Automated Reasoning* **53**, 141–172 (2014), <https://doi.org/10.1007/s10817-014-9301-5>
12. Brown, C.: Improved projection for cylindrical algebraic decomposition. *Journal of Symbolic Computation* **32**(5), 447–465 (2001), <https://doi.org/10.1006/jsc.2001.0463>
13. Brown, C.: QEPCAD B: A program for computing with semi-algebraic sets using CADs. *ACM SIGSAM Bulletin* **37**(4), 97–108 (2003), <https://doi.org/10.1145/968708.968710>
14. Brown, C., Davenport, J.: The complexity of quantifier elimination and cylindrical algebraic decomposition. In: *Proc. 32nd International Symposium on Symbolic and Algebraic Computation*. pp. 54–60. ISSAC '07, ACM (2007), <https://doi.org/10.1145/1277548.1277557>
15. Brown, C., Kosta, M.: Constructing a single cell in cylindrical algebraic decomposition. *Journal of Symbolic Computation* **70**, 14 – 48 (2015), <https://doi.org/10.1016/j.jsc.2014.09.024>
16. Caviness, B., Johnson, J.: *Quantifier Elimination and Cylindrical Algebraic Decomposition*. Texts & Monographs in Symbolic Computation, Springer-Verlag (1998), <https://doi.org/10.1007/978-3-7091-9459-1>
17. Chen, C., Moreno Maza, M.: An incremental algorithm for computing cylindrical algebraic decompositions. In: Feng, R., Lee, W., Sato, Y. (eds.) *Computer Mathematics*, pp. 199–221. Springer Berlin Heidelberg (2014), https://doi.org/10.1007/978-3-662-43799-5_17
18. Chen, C., Moreno Maza, M., Xia, B., Yang, L.: Computing cylindrical algebraic decomposition via triangular decomposition. In: *Proc. 34th International Symposium on Symbolic and Algebraic Computation*. pp. 95–102. ISSAC '09, ACM (2009), <https://doi.org/10.1145/1576702.1576718>
19. Collins, G.: Quantifier elimination for real closed fields by cylindrical algebraic decomposition. In: *Proc. 2nd GI Conference on Automata Theory and Formal Languages*. pp. 134–183. Springer-Verlag (reprinted in the collection [16]) (1975), https://doi.org/10.1007/3-540-07407-4_17
20. Collins, G., Hong, H.: Partial cylindrical algebraic decomposition for quantifier elimination. *Journal of Symbolic Computation* **12**, 299–328 (1991), [https://doi.org/10.1016/S0747-7171\(08\)80152-6](https://doi.org/10.1016/S0747-7171(08)80152-6)
21. Davenport, J., Bradford, R., England, M., Wilson, D.: Program verification in the presence of complex numbers, functions with branch cuts etc. In: *14th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing*. pp. 83–88. SYNASC '12, IEEE (2012), <http://dx.doi.org/10.1109/SYNASC.2012.68>
22. Davenport, J., Heintz, J.: Real quantifier elimination is doubly exponential. *Journal of Symbolic Computation* **5**(1-2), 29–35 (1988), [https://doi.org/10.1016/S0747-7171\(88\)80004-X](https://doi.org/10.1016/S0747-7171(88)80004-X)
23. Dolzmann, A., Seidl, A., Sturm, T.: Efficient projection orders for CAD. In: *Proc. 29th International Symposium on Symbolic and Algebraic Computation*. pp. 111–118. ISSAC '04, ACM (2004), <https://doi.org/10.1145/1005285.1005303>
24. Dolzmann, A., Sturm, T.: REDLOG: Computer algebra meets computer logic. *SIGSAM Bull.* **31**(2), 2–9 (1997), <https://doi.org/10.1145/261320.261324>

25. England, M.: Machine learning for mathematical software. In: Davenport, J., Kauers, M., Labahn, G., Urban, J. (eds.) *Mathematical Software – Proc. ICMS 2018*. Lecture Notes in Computer Science, vol. 10931, pp. 165–174. Springer International Publishing (2018), https://doi.org/10.1007/978-3-319-96418-8_20
26. England, M., Bradford, R., Chen, C., Davenport, J., Moreno Maza, M., Wilson, D.: Problem formulation for truth-table invariant cylindrical algebraic decomposition by incremental triangular decomposition. In: Watt, S., Davenport, J., Sexton, A., Sojka, P., Urban, J. (eds.) *Intelligent Computer Mathematics, Lecture Notes in Artificial Intelligence*, vol. 8543, pp. 45–60. Springer International (2014), http://dx.doi.org/10.1007/978-3-319-08434-3_5
27. England, M., Bradford, R., Davenport, J.: Improving the use of equational constraints in cylindrical algebraic decomposition. In: *Proc. 40th International Symposium on Symbolic and Algebraic Computation*. pp. 165–172. ISSAC '15, ACM (2015), <http://dx.doi.org/10.1145/2755996.2756678>
28. England, M., Bradford, R., Davenport, J., Wilson, D.: Choosing a variable ordering for truth-table invariant cylindrical algebraic decomposition by incremental triangular decomposition. In: Hong, H., Yap, C. (eds.) *Mathematical Software – ICMS 2014*. Lecture Notes in Computer Science, vol. 8592, pp. 450–457. Springer Heidelberg (2014), http://dx.doi.org/10.1007/978-3-662-44199-2_68
29. England, M., Errami, H., Grigoriev, D., Radulescu, O., Sturm, T., Weber, A.: Symbolic versus numerical computation and visualization of parameter regions for multistationarity of biological networks. In: Gerdt, V., Koepf, W., Seiler, W., Vorozhtsov, E. (eds.) *Computer Algebra in Scientific Computing (CASC)*, Lecture Notes in Computer Science, vol. 10490, pp. 93–108. Springer International Publishing (2017), https://doi.org/10.1007/978-3-319-66320-3_8
30. England, M., Wilson, D., Bradford, R., Davenport, J.: Using the Regular Chains Library to build cylindrical algebraic decompositions by projecting and lifting. In: Hong, H., Yap, C. (eds.) *Mathematical Software – ICMS 2014*. Lecture Notes in Computer Science, vol. 8592, pp. 458–465. Springer Heidelberg (2014), http://dx.doi.org/10.1007/978-3-662-44199-2_69
31. Erascu, M., Hong, H.: Real quantifier elimination for the synthesis of optimal numerical algorithms (Case study: Square root computation). *Journal of Symbolic Computation* **75**, 110–126 (2016), <https://doi.org/10.1016/j.jsc.2015.11.010>
32. Graebe, H., Nareike, A., Johanning, S.: The SymbolicData project: Towards a computer algebra social network. In: England, M., Davenport, J., Kohlhase, A., Kohlhase, M., Libbrecht, P., Neuper, W., Quaresma, P., Sexton, A., Sojka, P., Urban, J., Watt, S. (eds.) *Joint Proceedings of the MathUI, OpenMath and ThEdu Workshops and Work in Progress track at CICM*. *CEUR Workshop Proceedings* **1186**, (2014), <http://ceur-ws.org/Vol-1186/#paper-21>
33. Heinle, A., Levandovskyy, V.: The SDEval benchmarking toolkit. *ACM Communications in Computer Algebra* **49**(1), 1–9 (2015), <https://doi.org/10.1145/2768577.2768578>
34. Hong, H.: An improvement of the projection operator in cylindrical algebraic decomposition. In: *Proc. 15th International Symposium on Symbolic and Algebraic Computation*. pp. 261–264. ISSAC '90, ACM (1990), <https://doi.org/10.1145/96877.96943>
35. Huang, Z., England, M., Davenport, J., Paulson, L.: Using machine learning to decide when to precondition cylindrical algebraic decomposition with Groebner bases. In: *18th International Symposium on Symbolic and Numeric Algorithms for*

- Scientific Computing (SYNASC '16). pp. 45–52. IEEE (2016), <https://doi.org/10.1109/SYNASC.2016.020>
36. Huang, Z., England, M., Wilson, D., Bridge, J., Davenport, J.H., Paulson, L.: Using machine learning to improve cylindrical algebraic decomposition. *Mathematics in Computer Science*, volume to be assigned, Springer, 2019. <http://dx.doi.org/10.1007/s11786-019-00394-8>
 37. Huang, Z., England, M., Wilson, D., Davenport, J., Paulson, L., Bridge, J.: Applying machine learning to the problem of choosing a heuristic to select the variable ordering for cylindrical algebraic decomposition. In: Watt, S., Davenport, J., Sexton, A., Sojka, P., Urban, J. (eds.) *Intelligent Computer Mathematics, Lecture Notes in Artificial Intelligence*, vol. 8543, pp. 92–107. Springer International (2014), http://dx.doi.org/10.1007/978-3-319-08434-3_8
 38. Iwane, H., Yanami, H., Anai, H., Yokoyama, K.: An effective implementation of a symbolic-numeric cylindrical algebraic decomposition for quantifier elimination. In: *Proc. 2009 conference on Symbolic Numeric Computation*. pp. 55–64. SNC '09 (2009), <https://doi.org/10.1145/1577190.1577203>
 39. Jovanovic, D., de Moura, L.: Solving non-linear arithmetic. In: Gramlich, B., Miller, D., Sattler, U. (eds.) *Automated Reasoning: 6th International Joint Conference (IJCAR)*, *Lecture Notes in Computer Science*, vol. 7364, pp. 339–354. Springer (2012), https://doi.org/10.1007/978-3-642-31365-3_27
 40. Kobayashi, M., Iwane, H., Matsuzaki, T., Anai, H.: Efficient subformula orders for real quantifier elimination of non-prenex formulas. In: Kotsireas, S., Rump, M., Yap, K. (eds.) *Mathematical Aspects of Computer and Information Sciences (MACIS '15)*. *Lecture Notes in Computer Science*, vol. 9582, pp. 236–251. Springer International Publishing (2016), https://doi.org/10.1007/978-3-319-32859-1_21
 41. Kühlwein, D., Blanchette, J., Kaliszyk, C., Urban, J.: MaSh: Machine learning for sledgehammer. In: Blazy, S., Paulin-Mohring, C., Pichardie, D. (eds.) *Interactive Theorem Proving, Lecture Notes in Computer Science*, vol. 7998, pp. 35–50. Springer Berlin Heidelberg (2013), https://doi.org/10.1007/978-3-642-39634-2_6
 42. Liang, J., Hari Govind, V., Poupart, P., Czarnecki, K., Ganesh, V.: An empirical study of branching heuristics through the lens of global learning rate. In: Gaspers, S., Walsh, T. (eds.) *Theory and Applications of Satisfiability Testing – SAT 2017*, *Lecture Notes in Computer Science*, vol. 10491, pp. 119–135. Springer International Publishing (2017), https://doi.org/10.1007/978-3-319-66263-3_8
 43. McCallum, S.: An improved projection operation for cylindrical algebraic decomposition. In: Caviness, B., Johnson, J. (eds.) *Quantifier Elimination and Cylindrical Algebraic Decomposition*, pp. 242–268. *Texts & Monographs in Symbolic Computation*, Springer-Verlag (1998), https://doi.org/10.1007/978-3-7091-9459-1_12
 44. McCallum, S., Parusiński, A., Paunescu, L.: Validity proof of Lazard’s method for CAD construction. *Journal of Symbolic Computation* **92**, 52–69 (2019), <https://doi.org/10.1016/j.jsc.2017.12.002>
 45. Mulligan, C., Bradford, R., Davenport, J., England, M., Tonks, Z.: Non-linear real arithmetic benchmarks derived from automated reasoning in economics. In: Bigatti, A., Brain, M. (eds.) *Proc. 3rd Workshop on Satisfiability Checking and Symbolic Computation (SC² 2018)*. pp. 48–60. *CEUR Workshop Proceedings* **2189**, (2018), <http://ceur-ws.org/Vol-2189/>
 46. Mulligan, C., Davenport, J., England, M.: TheoryGuru: A Mathematica package to apply quantifier elimination technology to economics. In: Davenport, J., Kauers, M., Labahn, G., Urban, J. (eds.) *Mathematical Software – Proc. ICMS 2018*. *Lecture Notes in Computer Science*, vol. 10931, pp. 369–378. Springer International Publishing (2018), https://doi.org/10.1007/978-3-319-96418-8_44

47. Paulson, L.: Metitarski: Past and future. In: Beringer, L., Felty, A. (eds.) *Interactive Theorem Proving*, Lecture Notes in Computer Science, vol. 7406, pp. 1–10. Springer (2012), https://doi.org/10.1007/978-3-642-32347-8_1
48. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830 (2011), <http://www.jmlr.org/papers/v12/pedregosa11a.html>
49. Platzer, A., Quesel, J., Rümmer, P.: Real world verification. In: Schmidt, R. (ed.) *Automated Deduction (CADE-22)*, Lecture Notes in Computer Science, vol. 5663, pp. 485–501. Springer Berlin Heidelberg (2009), https://doi.org/10.1007/978-3-642-02959-2_35
50. Strzeboński, A.: Cylindrical algebraic decomposition using validated numerics. *Journal of Symbolic Computation* **41**(9), 1021–1038 (2006), <https://doi.org/10.1016/j.jsc.2006.06.004>
51. Strzeboński, A.: Cylindrical algebraic decomposition using local projections. *Journal of Symbolic Computation* **76**, 36–64 (2016), <https://doi.org/10.1016/j.jsc.2015.11.018>
52. Sturm, T.: New domains for applied quantifier elimination. In: Ganzha, V., Mayr, E., Vorozhtsov, E. (eds.) *Computer Algebra in Scientific Computing*, Lecture Notes in Computer Science, vol. 4194, pp. 295–301. Springer Berlin Heidelberg (2006), https://doi.org/10.1007/11870814_25
53. Tarski, A.: *A Decision Method For Elementary Algebra And Geometry*. RAND Corporation, Santa Monica, CA (reprinted in the collection [16]) (1948)
54. Urban, J.: MaLAREa: A metasystem for automated reasoning in large theories. In: *Empirically Successful Automated Reasoning in Large Theories (ESARLT '07)*, *CEUR Workshop Proceedings*, **257**, pp. 14, (2007), <http://ceur-ws.org/Vol-257/>
55. Wilson, D., Bradford, R., Davenport, J., England, M.: Cylindrical algebraic sub-decompositions. *Mathematics in Computer Science* **8**, 263–288 (2014), <http://dx.doi.org/10.1007/s11786-014-0191-z>
56. Wilson, D., Davenport, J., England, M., Bradford, R.: A “piano movers” problem reformulated. In: *15th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing*. pp. 53–60. SYNASC '13, IEEE (2013), <http://dx.doi.org/10.1109/SYNASC.2013.14>
57. Wilson, D., England, M., Davenport, J., Bradford, R.: Using the distribution of cells by dimension in a cylindrical algebraic decomposition. In: *16th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing*. pp. 53–60. SYNASC '14, IEEE (2014), <http://dx.doi.org/10.1109/SYNASC.2014.15>
58. Wilson, D., Bradford, R., Davenport, J.: A repository for CAD examples. *ACM Communications in Computer Algebra* **46**(3), 67–69 (2012), <https://doi.org/10.1145/2429135.2429137>
59. Xu, L., Hutter, F., Hoos, H., Leyton-Brown, K.: SATzilla: Portfolio-based algorithm selection for SAT. *Journal Of Artificial Intelligence Research* **32**, 565–606 (2008), <https://doi.org/10.1613/jair.2490>